



Citation for published version:

Day, M 2003, *Prospects for institutional e-print repositories in the United Kingdom*. UKOLN, University of Bath.

Publication date:
2003

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Prospects for institutional e-print repositories in the United Kingdom

ePrints UK Supporting Study, no. 1 (May 2003)¹

Michael Day

UKOLN, University of Bath, Bath BA2 7AY, United Kingdom
m.day@ukoln.ac.uk

Abstract. This study introduces ePrints UK, a project funded as part of the JISC's Focus on Access to Institutional Resources (FAIR) Programme. It first introduces the project and the main features of the FAIR programme as it relates to e-print repositories. Then it provides some general information on open-access principles, institutional repositories and the technical developments that have made their development viable. There follows a review of relevant repositories in the UK and an indication of what impact ePrints UK might have in supporting learning, teaching and research. This is followed by a discussion of perceived impediments to the take-up of institutional repositories, including both practical and cultural issues. A final section investigates the development of ongoing evaluation criteria for the project.

Introduction

EPrints UK (<http://www.rdn.ac.uk/projects/eprints-uk/>) is a two-year project funded by the Joint Information Systems Committee (JISC) as part of its Focus on Access to Institutional Resources (FAIR) Programme. The aim of the project is to develop a national service provider repository of e-print records based at the University of Bath, derived by harvesting metadata from institutional and subject-based e-prints archives using the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). In addition, the project aims to provide access to these institutional assets through the eight Resource Discovery Network (RDN) faculty level hubs and the Education Portal based at the University of Leeds. It is also investigating the use of Web Services technologies for the enhancement of metadata and for the automatic linking of citations. In addition, the project is producing a series of four supporting studies on

¹ ePrints UK Supporting Study, no. 1, Version 1.0, first published online on 28 May 2003.
HTML version available from: <http://eprints-uk.rdn.ac.uk/project/docs/studies/impact/>

issues relating to the creation, maintenance and sustainability of institutional e-print repositories.

This report is the first of these studies and will first introduce the concept of institutional e-print repositories in the context of open-access initiatives. It will introduce ePrints UK and other projects funded by the FAIR Programme, and outline how the ePrints UK project might have an impact in supporting learning, teaching and research activity in the UK higher and further education sectors. This will be followed by a discussion of some of the possible reasons why institutional repositories may not succeed and some criteria for ongoing evaluation.

2. The project context

2.1 ePrints UK

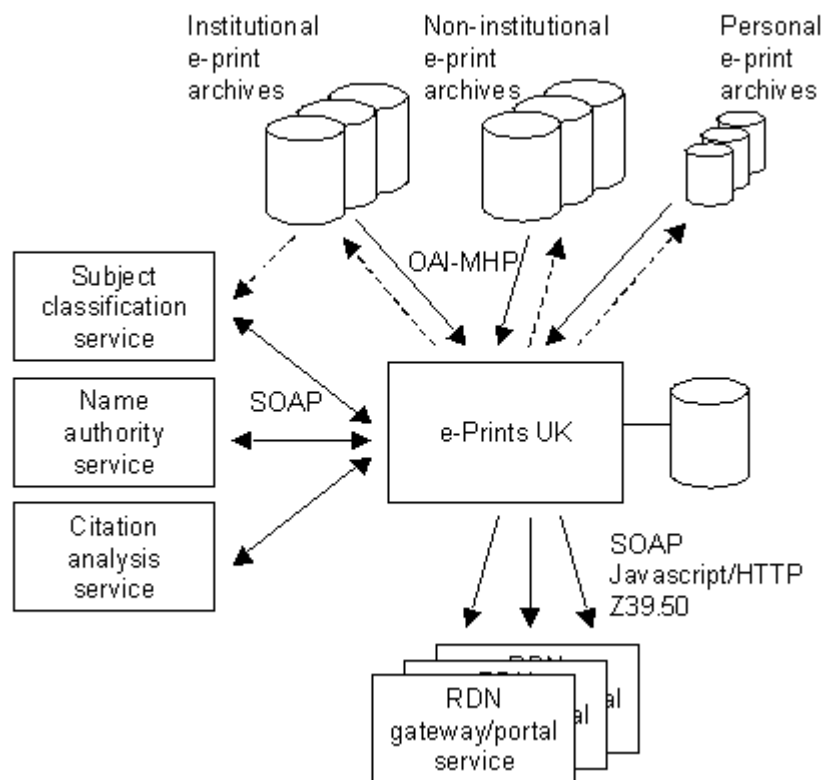
The ePrints UK project is a collaborative initiative of UKOLN, the Resource Discovery Network (RDN), the University of Southampton and OCLC Research. Its primary objective is to develop a national service through which the UK higher and further education communities can access the collective output of e-prints from UK repositories. The architecture (Figure 1) of this service will be based on harvesting metadata from OAI-PMH compliant e-print repositories at UK institutions into a centralised database

Once gathered, both the metadata and the full-text of e-prints (where this is available) will then be passed to external Web Services that will be able to enhance metadata records by:

- Adding (or validating) authoritative forms of author names
- Automatically assigning subject classification terms
- Parsing bibliographic references into structured forms, using the OpenURL standard

The enhanced metadata will then form the basis of an ePrints UK service, which will be made available to end-users in a number of different ways. Firstly, through a general search interface, integrated into the RDN's Web pages, which will provide access to all of the harvested and enhanced metadata. Secondly, through the development of configurable discovery services that would enable RDN hubs, academic institutions and other organisations to embed ePrints UK directly within their own services. The project will test this approach by embedding ePrints UK into the eight subject-based RDN 'hub' services.

It is hoped that ePrints UK will help support the adoption of e-prints within the UK HE and FE communities by giving academics an incentive to deposit papers in institutional repositories and other e-print services. The project is hosting a number of events that will encourage institutions to make e-prints available, and to facilitate common approaches to metadata practice.

Figure 1 - ePrints UK architecture

(Andy Powell, UKOLN)

2.2 The FAIR Programme

EPrints UK is just one of a series of projects that make up the FAIR programme. The programme focuses on three main areas: e-prints and electronic theses, museums and images, and institutional portals. As part of FAIR, the JISC funded six projects with a particular focus on e-prints and two on electronic theses. Apart from ePrints UK, the FAIR projects that mostly concern e-print type resources are:

- DAEDALUS - a project concerned with the establishment of a range of OAI-PMH-compliant digital collections at the University of Glasgow. These will include e-prints, theses, finding aids and institutional documents (<http://www.lib.gla.ac.uk/daedalus/>)
- Electronic Theses - a project led by Robert Gordon University evaluating a wide range of existing models of e-theses production and access (<http://www2.rgu.ac.uk/library/e-theses.htm>)
- HaIRST (Harvesting Institutional Resources in Scotland Testbed) - a project led by the University of Strathclyde concerned with the

development of a pilot service that will provide access to institutional resources in Scotland, including e-prints (<http://hairst.cdli.strath.ac.uk/>)

- RoMEO (Rights Metadata for Open archiving) - a project undertaken by Loughborough University investigating the rights issues that surround the 'self-archiving' of research in the UK academic community and the recommendation of a rights metadata scheme (<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>)
- SHERPA (Securing a Hybrid Environment for Research Preservation and Access) - a project that will create a number of OAI-PMH-compliant e-print repositories and which will investigate key issues in the creation and maintenance of e-print collections. These issues will include intellectual property rights (IPR), quality control, collection development, business models, scholarly communication cultures, and institutional strategies. The project is supported by the Consortium of University Research Libraries (CURL) and is led by the University of Nottingham (<http://www.sherpa.ac.uk/>)
- TARDIS - a project that is developing an institutional e-print repository for the University of Southampton (<http://tardis.eprints.org/>)
- Theses Alive! - a project led by the University of Edinburgh examining the use of OAI-compliant software for electronic theses.

Projects, therefore, are involved in both fostering the creation of institutional e-print repositories within the UK academic community (e.g., DAEDALUS, SHERPA, TARDIS) and the development of new types of services that provide some kind of unified access to the content of such repositories (e.g., ePrints UK, HaIRST). Many projects are also investigating a range of technical and non-technical issues relating to the creation, maintenance and sustainability of institutional e-print repositories (e.g., ePrints UK, RoMEO, SHERPA).

3. The wider context

Institutional e-print repositories are one of a range of responses to what is generally known as the serials pricing crisis. The core of this crisis is that journal subscription prices have been rising rapidly while the budgets of those institutions that subscribe remain stable or are in decline. As a result, subscriptions are cancelled, and the subscription prices rise even further, resulting in a vicious circle of rising prices and further cancellations. One suggested way out of the serials pricing crisis is the adoption of open-access principles. This section will introduce open-access, define institutional repositories and outline some of the technical developments that have made them possible.

3.1 The open-access movement

The costs of the present journal-based communication system, both in terms of subscription prices and restricted access, have meant that the impending demise of

traditional journals has been confidently predicted for a number of years (e.g., LaPorte, et al., 1995; Odlyzko, 1995). However, for a variety of reasons, this has not yet happened. So, while most publishers have willingly embraced the Internet as an efficient distribution medium for journals, much content is still 'hidden' behind access barriers, typically paid for by the user (or the institution for which they work) through subscriptions, licences or by pay-per-view.

These practices seem to conflict with the aims of scientists and scholars who, it is argued, primarily publish research papers for research impact, i.e. in having their work read, cited and built on in the research of others (Harnad, 2001, p. 1024). For example, Gordon Fletcher (2002, p. 6) of BioMed Central has said that when scientists and clinicians submit papers, "they give away the fruits of their labour in order to advance scientific progress and to register their part in that advancement." In this context, any kind of cost barrier seems counter-intuitive. Thus Harnad (2001, p. 1024) says that from the authors' viewpoint, "toll-gating access is as counterproductive as toll-gating access to commercial advertisements." The existence of the Internet, however, provides authors with several ways of resolving these conflicting priorities. No longer, as in the print era, do authors need to trade the copyright of works to publishers in exchange for having them printed and distributed (Harnad & Hey, 1995, p.114). Instead, authors are free to publish in new generations of open-access journals or to deposit digital copies of research papers (e-prints) in publicly available repositories.

The Budapest Open Access Initiative (BOAI) supports these general aims, and stresses the 'public good' nature of providing unrestricted access to the peer-reviewed scientific literature (<http://www.soros.org/openaccess/>).

Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.

Those scientists and scholars that support the initiative are encouraged to facilitate its aims in two ways, firstly by supporting open-access journals, secondly through the 'self-archiving' of peer-reviewed research papers (or e-prints).

Open-access journals include both those existing titles that have committed to providing some kind of free access to their content - e.g., through initiatives like PubMed Central (<http://www.pubmedcentral.nih.gov/>) - as well as a new generation of e-journals based on non-subscription business models. Good examples of the latter are the series of journals that are published by BioMed Central and the two new titles recently proposed by the Public Library of Science (PLOS) initiative (Science, 2002). These work on the general principle that the authors of research papers should pay for them to be published, while access to the journals should be free. The amount that would be paid varies from journal to journal. For example, BioMed Central currently charge \$500 per paper as an 'article processing charge,' while the PLOS initiative has estimated that authors will be asked to pay around \$1,500 per paper when their journals become available sometime in 2003.

The other proposed way to support open access is for authors to continue to publish in existing peer-reviewed journals but to supplement this by depositing copies of published papers in public e-print repositories. The basic model for these types of

service is arXiv.org, the physics, mathematics and computer science repository operated by Cornell University (<http://arxiv.org/>). This was set up at the Los Alamos National Laboratory (LANL) in the early 1990s to facilitate the communication requirements of high-energy physicists (Ginsparg, 1994). Now based at Cornell University, it remains one of the largest e-print repositories with (in March 2003) almost 230,000 papers, growing by more than 30,000 per year. Other subject-based e-print repositories exist, e.g. for cognitive sciences (CogPrints) or chemistry (the Elsevier-run Chemistry Preprint Server), but there is now a growing interest in the creation of interoperable repositories based on research or educational institutions.

3.2 Institutional repositories

Institutional repositories have been defined in a recent Scholarly Publishing and Academic Resources Coalition (SPARC) position paper as digital collections that capture and preserve the intellectual output of a single or multi-university community (Crow, 2002, p. 4). While repositories like arXiv or CogPrints focus on particular subject domains, institutional repositories store and make accessible the outputs of institutions. In this, they are not necessarily limited to e-prints of the research literature, but can provide an institutional focus for the collection and preservation of scientific data, learning resources, image collections and many other different types of content.

According to Crow (2002, p. 6), institutional repositories have two main rationales. Firstly, they will form part of a global system of distributed interoperable repositories that will help facilitate reform of the scholarly communication system. Secondly, they will help enhance an institution's visibility and prestige, "making it easier to demonstrate its scientific, social and financial value." To these can be added a third; the editor-in-chief of Nature's comment that institution-based repositories may help support accountability by hosting archives of the data produced in their laboratories. Campbell (2002, p. 964) has noted that, "increasing attention to the prevention of [scientific] misconduct requires institutions to keep better records of their researchers' practices at the lab bench."

It has been argued that the development of institutional repositories of e-prints will help facilitate open-access to the products of research. While some librarians may hope that the widespread adoption of open-access principles will (in time) help offset the adverse consequences of the ongoing 'serials crisis,' most of the scientific proponents of e-prints focus on the benefits of free and open access to the products of research. In this, they can point to a long-standing tradition within science for the sharing of data, e.g. through scientific data archives or public DNA sequence databases like EMBL-Bank. As with sequence databases, much has been made of the way that providing open-access to the unified scientific literature might encourage "the development of new, more sophisticated, and valuable ways of using this information" (Roberts, et al., 2001, p. 2318).

3.3 Technical developments

The existence of institutional repositories has been made possible by the development of standards and tools that facilitate interoperability between multiple repositories. The most important of these is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

The OAI-PMH was developed from the 'Santa Fe Convention' agreed at the initial meeting of the Open Archives Initiative (OAI) in 1999. Version 1.0 of the protocol was issued in January 2001; the latest version (Version 2.0) being released in June 2002 (<http://www.openarchives.org/>). The protocol provides an interoperability framework based on the harvesting of metadata. To that end, it defines a simple set of metadata that can facilitate federated resource discovery. The protocol divides users of the OAI-PMH into two main categories:

- *Data providers* - those who provide repository services and populate them with metadata
- *Service providers* - those that harvest this metadata to create services, chiefly to aid the discovery of the resources described by data providers but also as a basis for building value-added services.

In short, the OAI-PMH uses metadata harvesting to allow data providers to share metadata with service providers. Under this model, for example, the proposed national ePrints UK service would be an OAI service provider while the institutional repositories set up as part of DAEDALUS, SHERPA, TARDIS or other UK initiatives would be data providers. Existing OAI service providers include the experimental ARC service developed by Old Dominion University that federates access to the content of a large number of data providers (<http://arc.cs.odu.edu/>). While it is envisaged that data providers will normally use the OAI protocol to make metadata publicly available through the Internet, there has also been some use of the protocol to share metadata within 'closed' systems. An example of this is the use of the OAI-PMH by the Resource Discovery Network (RDN) to support the cross-search of Internet subject gateways (Powell, 2001).

The OAI-PMH defines a metadata element set based on simple Dublin Core. Its functionality has deliberately been kept very simple. So, for example, the specification includes just six protocol requests, including GetRecord and Identify, ListIdentifier, ListMetadataFormats, ListRecords. The OAI-PMH links these relatively simple components to allow distributed e-print archives to be combined into a single conceptual repository. OAI service providers can select content based on a range of criteria, e.g. subject matter, geographical location of repository, etc.

Version 1.0 of the OAI-PMH was released in January 2001 for experimental use, the latest (v. 2.0) in June 2002. A growing number of open-source software tools that help support the implementation of the protocol are now becoming available (e.g. Van de Sompel & Lagoze, 2002). For data providers, the most important of these are the University of Southampton's EPrints software and MIT Libraries's DSpace system.

The GNU EPrints software was developed to facilitate the deployment of OAI-PMH compliant repositories (data services) for research papers, although it can be configured to deal with other kinds of digital information. The software was developed by the Department of Electronics and Computer Science at the University of Southampton and is freely available from the EPrints.org Web site

(<http://software.eprints.org/>). Versions 1 and 2 of the software has (to date) been implemented by over sixty-five repositories, both subject-based (e.g. by CogPrints) and institutional (e.g., the universities of Glasgow, Nottingham and Munich)

A second software toolkit that can be used to develop institutional repositories is DSpace (<http://dspace.org/>), developed by the Massachusetts Institute of Technology (MIT Libraries) in collaboration with Hewlett-Packard (HP Labs). DSpace has more ambitious objectives than the EPrints software, aiming to provide repositories of a wide range of institutional outputs, including research papers, books, theses, data sets, programs, multimedia, etc. As the backbone of a repository, DSpace provides a way for institutions to manage these outputs, make them more widely available, and ultimately to preserve them (Smith, 2002; Smith, et al., 2003). In order to support interoperability with other digital repositories, DSpace has implemented the OAI-PMH. DSpace has been implemented at MIT and is being tested elsewhere, e.g. through the DSpace@Cambridge project based at the University of Cambridge (<http://www.lib.cam.ac.uk/dspace/>).

4. Institutional e-print repositories in the UK

The success of ePrints UK and some of the other FAIR Programme projects will depend upon the widespread availability of suitable content. In terms of the OAI, the proposed national ePrints UK service is primarily a Service Provider. Therefore, in order for its innovative use of Web Services for metadata enhancement or citation linking to be seen as useful, it will have to demonstrate this on a significant body of appropriate content. The following sections will assess the current situation with regard to institutional e-print repositories in the UK and look at some potential impediments to success.

While a number of subject-based e-print repositories are based (or have mirror sites) in the UK, there are, to date, very few institutional repositories. Those that do exist, contain a relatively small number of e-prints. For example, in May 2003, the list of known services running the EPrints software included just nine repositories based within the ac.uk domain, most of which contained relatively few records (Table 1). The largest repository (from the Department of Electronics and Computer Science at the University of Southampton) contained over seven thousand records but all of the others could only collectively muster 176. It should also be noted that not all of these will include full-text.

Pinfield (2003) in a recent D-Lib Magazine article on the development of open archives in the UK has argued that more effort needs now be put into actually populating repositories.

Setting up an institutional repository and designing collection management policies are relatively straightforward; populating the repository is not. The content of institutional repositories needs to come largely from researchers within the institution, and persuading them to submit this content is a major challenge. Self-archiving requires a cultural change amongst researchers that can only be achieved through significant advocacy activity, and even then it will probably happen only gradually.

Some of the other FAIR Programme projects (e.g., DAED AUS, SHERPA and TARDis) are attempting to foster the take-up of repositories within particular institutions (or groups of institutions) but Pinfield's comments suggest that a wider advocacy role may be required. This would need both to focus on the incentives for academics and researchers to deposit in institutional repositories and to answer some of their concerns.

Table 1 - Repositories in the ac.uk domain using EPrints software, March 2003

Institutional repositories	Records	URL
University of Southampton, Department of Electronics and Computer Science	7,158	http://eprints.ecs.soton.ac.uk/
University of Edinburgh, Theoretical and Applied Linguistics	90*	http://archive.ling.ed.ac.uk/
University of Nottingham (Nottingham ePrints)	46	http://www-db.library.nottingham.ac.uk/ep1/
University of Glasgow (ePrints @ Glasgow)	36	http://eprints.lib.gla.ac.uk/
University of Bath [pilot]	4	http://eprints.bath.ac.uk/
Strathclyde University (StrathPrints)	0	http://eprints.cdrl.strath.ac.uk/
Subject-based repositories	Records	URL
CogPrints Cognitive Science Eprint Archive	1,709	http://cogprints.ecs.soton.ac.uk/
Psychology [e-journal]	720	http://psycprints.ecs.soton.ac.uk/
Formations Media Studies Archive	21	http://formations2.ulst.ac.uk/

Sources: <http://software.eprints.org/>; * from search of: <http://archive.ling.ed.ac.uk/> (28 May 2003)

4.1 The potential impact of ePrints UK

It is difficult to evaluate the potential impact of ePrints UK without noting that, at the present time, there are a limited number of institutional repositories in the UK. Pinfield is also correct in concluding that more consideration needs to be given to populating those repositories that do exist. The challenge of the FAIR programme and related initiatives will be to help foster the setting up of well-populated institutional repositories in UK higher and further education institutions.

Van de Sompel and Lagoze (2002, p. 145) note that technologies like the OAI-PMH exhibit network effects, in that "initial adoption may be slow and steady and positive feedback then dramatically increases the adoption rate." This means that can be difficult to measure success, particularly in the early stages of adoption. The same will apply to measuring the long-term impact of institutional repositories.

In the shorter term, the creation of institutional repositories of e-prints will affect stakeholders in different ways. Institutions themselves can use repositories as a

symbol of their commitment to the sustainable management of its own intellectual resources. As Lynch (2003) has said:

At the most basic and fundamental level, an institutional repository is a recognition that the intellectual life and scholarship of our universities will increasingly be represented, documented, and shared in digital form, and that a primary responsibility of our universities is to exercise stewardship over these riches: both to make them available and to preserve them. An institutional repository is the means by which our universities will address this responsibility both to the members of their communities and to the public. It is a new channel for structuring the university's contribution to the broader world, and as such invites policy and cultural reassessment of this relationship.

Institutional repositories can also help raise the profile of the institution. Within the institution, there are potential opportunities for university libraries to move beyond their traditional custodial roles (although these will remain important) so that they can contribute actively to the development of the scholarly communication process itself (Crow, 2002, p. 20).

There are also many potential benefits for individual academics or researchers, especially if the number and size of repositories reach some kind of 'critical-mass.' A large number of research papers are already being made available through academics' personal home pages, the Web pages of research projects, laboratories or departments, etc. Depositing these in institutional repositories will produce additional ways of finding this material, and to help ensure continued access when researchers change jobs or when institutions reorganise. Crow (2002, p. 23) comments that the principal benefits for authors are an enhanced professional visibility, and from the increased article impact that open-access makes possible. Lawrence (2001) found that, at least for computer science conference papers, there was a "clear correlation between the number of times an article is cited and the probability that the article is online." Students may also benefit because the research output of their institution (and others) can be made accessible through virtual learning environments or the library catalogue.

The potential impact of the ePrints UK project itself is, of necessity, more limited. Its twin national and subject-based approaches will (hopefully) demonstrate to UK universities and colleges the benefits of setting up institutional repositories using the OAI-PMH. The Web services for supporting subject classification and name authorities will potentially have other uses in digital libraries, e.g. as tools to support metadata creation. The citation analysis tools being developed at the University of Southampton may eventually offer the opportunity of developing advanced scientometric tools in support of research evaluation.

The focus on e-prints will mean that the project outputs will mostly support the research activities of UK universities and colleges, providing an additional (and sustainable) method of dissemination and - in a network of repositories - rapid access to the research output of other institutions. They will also help support learning and teaching, to the extent that they can help provide managed access to relevant research papers. In addition, institutional repositories may be able to contribute to the creation or management of online reading lists.

4.2 Evaluation criteria for ePrints UK

The ePrints UK project has a number of different technical outputs (e.g., Martin, 2003). Firstly, the project is developing a central database of e-print metadata using the ARC harvesting software. While the project proposal suggested that this would primarily provide access to UK repositories, this service will now be harvesting metadata (and full text where available) of e-print records from OAI compliant repositories in the UK and elsewhere. This service will need to be evaluated with regard to its perceived usefulness to users and as a demonstration of the value of institutional repositories.

A second technical output will be developing SOAP (Simple Object Access Protocol) interfaces to pass the metadata (and full text) to external Web services for the enhancement, augmentation, or validation of metadata. Two of these Web services have been developed by OCLC Research, and will be used to automatically assign:

- Dewey Decimal Classification (DDC) subject classification codes
- Standardised forms of name from authority files

Both of these services are experimental, and their outputs will need to be tested for their validity and usefulness. For example, it is likely that many of the authors of research papers will not be found in name authorities (like the Library of Congress Name Authority File) that are based on other types of publications, chiefly books. A third Web service, a citation analysis service, will parse semi-structured citation information in the document text to form structured, machine-readable, citations in the form of OpenURLs. The project plans to undertake an evaluation of the subject classification Web service, taking a statistical survey approach that analyses the DDC notation assigned to 400 metadata records.

The ePrints UK service will be made available to end users in a number of ways. Firstly, the project's main Web site will provide a search interface to all the enhanced, harvested metadata. In addition, ePrints UK will offer shared, configurable discovery services that enable the RDN hubs, UK academic institutions and other organisations to simply embed ePrints UK within their services, closely based on the RDN's existing RDN-Include and RDNi-Lite offerings. These services need to demonstrate the value to institutions of setting up repositories. However, this will be dependent on the existence of sufficient content, which Table 1 suggests may be problematic. Direct advocacy of institutional repositories is not part of the main scope of ePrints UK (although it is part of the FAIR programme itself or projects like SHERPA), so it would seem unfair to evaluate the ePrints UK service solely on this basis. As part of ePrints UK, RDN hubs will be organising small focus groups that will help evaluate the services offered by the project.

5. Potential impediments to success

There are probably many reasons why self-archiving has not (to date) succeeded in the UK. It could be related to a lack of awareness or opportunity, although the low

numbers of records in those institutional repositories that exist suggest that other factors may be at work. Pinfield (2003) is correct in surmising that self-archiving will require cultural change, but there are also a number of more practical issues that will need to be addressed. This section will first deal with a number of these issues - copyright, peer-review, preservation and economics - before going on to consider some of the wider cultural issues in more detail.

5.1 Practical issues

There are doubtless a number of reasons why e-prints and open-access principles have not yet caught the imagination of those who would be perceived to most benefit from them, i.e. individual researchers. While many academics and researchers seem happy to make copies of selected research papers available on institutional, project or personal Web pages, they appear to be less sure about the role of e-print repositories, whether subject-based or institutional. Concerns are often raised about practical issues like copyright or peer-review. This section will introduce some of these proposed impediments, and briefly consider their validity.

5.1.1 Copyright

One possible impediment to the success of institutional e-print repositories is the traditional assignment of copyright to publishers. In most cases when a paper has been accepted for publication in a journal, the author/s then assign the copyright to the publisher or (sometimes) grants them an 'exclusive license' to publish. In many cases, these contracts expressly forbid the publication of papers in any other form, including their deposit into e-print repositories. For example, the latest copyright agreement issued by Nature Publishing Group asks authors to grant them an exclusive licence to publish. Authors are allowed to "re-use the papers in any printed volume of which they are an author; to post a PDF copy on their own (not-for-profit) website; to copy (and for their institutions to copy) their papers for use in coursework teaching; and to re-use figures and tables" (Nature, 2003). However, the licence expressly excludes "open archival websites, such as those that host collections of articles by an institution's researchers." While in many cases it is possible to make changes to these contracts, many authors simply agree to the default terms. From a survey of authors and publishers, Gadd, Oppenheim & Proberts (2003) found that around a third of academics were not sure who owned the copyrights in a research paper. The same study showed that while 41 per cent of the surveyed academics freely assigned copyright to publishers, almost half (49%) did so reluctantly. As, Bide (2002, p. 24) comments, the "pressure on academic authors to publish (and to publish in high profile journals) may lead them to sign agreements that they may otherwise might not."

Those responsible for institutional repositories will have to be aware their responsibilities as de facto publishers. With regard to e-print repositories generally, Bide (2002, pp. 25-26) suggests that there will be a need for explicit agreements with depositing authors, maybe as an automated part of the submission process.

These should include (for example) warranties on the part of the author that they are not breaching any third party agreement - or copyright - by posting

the eprint. This would also ensure that authors explicitly accepted the terms under which the content is being made available to others.

It has been proposed that one way of solving at least some of the copyright issues of institutional repositories would be for universities and other educational institutions to assert copyright ownership of the research outputs of employees (Gadd, Oppenheim & Proberts, 2003). The JISC Scholarly Communications Group (2002) in their submission to the Research Support Libraries Group (RSLG) noted that an awareness of the negative consequences of copyright transfer to publishers had led some UK universities to reconsider their policies on copyright. Bide (2002, p. 23) has described the question of ownership of intellectual property rights of academics as "one of the more contentious issues" facing higher education. He notes that the terms of the Copyright, Designs and Patents Act, 1988 would normally mean that copyright made "in the course of employment" would pass to the employer. However, in practice, "most academic institutions do not exercise this right with respect to copyrights in journal articles or in textbooks." So, while many universities are now beginning to revisit their position with regard to the IPR of patents or pedagogic material, most policies still exclude academic articles or books.

5.1.2 Peer-review and quality control

Another objection to e-print repositories is that it might enable the bypassing of peer-review. Review is an essential part of the existing scientific and scholarly publishing process, especially in disciplines like medicine or chemistry. Peer-review, however, is outside the scope of the repository itself. The focus of an institutional repository can be on content that is either peer-reviewed or not, the choice being left to those who develop their collection policies. In order to ensure a certain level of quality control, some institutions may decide to separate peer-reviewed e-prints from those that have not been reviewed. The importance of this will vary between subject disciplines.

5.1.3 Long-term preservation

Another potential problem is what will happen to e-print repositories in the longer-term (e.g., Smith, 2003). Academics build on the work of others and will regularly cite or make reference to the past literature. Ziman (1968, p. 103) notes that the "citation of references validates many of the claims that ... [the scientist] will make in his [or her] paper and embeds it in the pre-existing consensus." One of the roles that printed journals have evolved to fulfil included the establishment of a public domain archive (Rowland, 1997). This was not seen as the responsibility of publishers, but was effectively a by-product of publication in printed form and it worked because research libraries collectively acted as a distributed repository, preserving the knowledge embodied in journals for current and future scholars.

The move towards licensing content threatens the role of libraries as the preserver of scientific knowledge. This brings both new threats and new opportunities. For example, the threats might include the possibility that institutions that set up repositories may not always be aware of their responsibility to ensure the long-term preservation of content. Even when they are, they may not have the organisational infrastructure or technical knowledge to do this content successfully. Other problems could arise when institutions restructure, merge or disappear. On the other hand, a key

opportunity is that educational and research institutions are often better placed to solve digital preservation challenges than are publishers.

5.2 Cultural issues

In addition to the practical issues considered above, there are many other issues that arise from the culture of scientific and scholarly endeavour. In particular, we need to focus on the multiple functions that peer-reviewed journals currently play in the scholarly communication system. While some advocates of e-prints argue that the authors of peer-reviewed papers write primarily for research impact (e.g., Harnad, 2001, p. 1024), the multiple roles that journals have evolved over time to fulfil suggest that this may not be the whole story. In any case, Schaffner (1994, p. 245) has noted that enabling technologies may not be, by themselves, sufficient to bring about major changes in communication forms. Odlyzko (1995, p. 86) reflects that "while scholars may be intellectually adventurous, they tend to be conservative in their work habits."

5.2.1 The role of journals in the current communication system

E-print repositories and open access journals are intended to help address some of the problems resulting from the serials pricing crisis and its successor, the permissions crisis (e.g., Suber, 2003). In this it is assumed that scientists and scholars produce peer-reviewed papers primarily for research impact, in having their work read, cited and built on in the research of others. This suggests that the main function of journals is dissemination. However, it is possible that this view of the scholarly communication process is incomplete.

In the first place, peer-reviewed journals are not, and have never been, the only way for scientists or scholars to disseminate. It is one of a range of different dissemination methods - including informal discussion, conference papers, pre-prints, books, etc. - that are now being supplemented in the digital era. Secondly, while dissemination remains one of the more important roles of peer-reviewed journals, they have evolved into a sophisticated system that provides (at least) the following additional features (Rowland, 1997):

- Quality control through editorial processes and peer-review
- The basis for the establishment of priority over advances in science and the recognition of authors
- The establishment of a distributed public-domain archive

In short, the multiple essential functions that are fulfilled by journals may mean that scientists and scholars may be reluctant to adopt forms of scientific communication that emphasise the importance of dissemination over its other roles. It is perhaps instructive that many of the papers deposited in arXiv are also submitted for publication in peer-reviewed journals, thus combining the rapid dissemination abilities of digital technology with the other functions best provided by journals. Following this, the advocates of self-archiving now argue that depositing e-prints in

institutional repositories need not mean that authors should give up publishing in high-impact journals. Thus Suber (2003) says that the open-access movement does not "call on scholars to shun priced or printed journals, either as authors, editors, referees, subscribers, or readers, nor do we call on libraries to cancel or deaccession them." However the move towards open-access is often based on a assumption that once alternative structures are in place, the true cost of the journal system will become apparent to university administrators (and others) and will result in the rapid decline of the existing journal system. Thus Odlyzko (1995, p. 87):

The problem is that the natural development of present preprint distribution systems ... is going to make scholarly papers freely available on the Net, so that scholars will be relying on their libraries less and less. They will therefore have less and less incentive to press for paper journal subscriptions to be maintained, which will lead to diminished circulation, and therefore to higher prices and more pressure from libraries to cut back on subscriptions.

In addition, the nature of scholarly communication means that there remain 'perverse incentives' for scholars and scientists to publish papers in the existing journal-based system (Odlyzko, 1997). Researchers and scholars choose whether (or not) to publish in journals depending on a range of objective or subjective criteria - e.g., prestige, perceived quality, audience, high-impact, etc. - but not primarily on price. Odlyzko (1997) notes that often the "incentives are to publish in high-cost outlets." The market for journals can be further skewed by the fact that the parts of institutions that actually spend money on journal subscriptions or licenses are not always the ones who read or submit articles to serials. Also, the use of document supply services may additionally insulate researchers and academics from some of the consequences of the serials crisis.

5.2.2 Different motives for writing and publishing papers

Another consequence of the assumption that dissemination is the primary objective of scholarly communication is that it ignores the many reasons why authors write and submit papers. So, for example, the establishment of priority over a particular advance or discovery is one of the basic motivations of most scientists and is, on occasion, considered more important than being read or cited by peers (Meadows, 1991). Close (1992, p. 299) has noted how closely priority is linked with publication.

Usually in science there is a great pressure to be first, to win the race and gain the honour of discovery. That honour requires acceptance by the community of science which in turn needs refereed publication of all the details necessary for the successful replication of the discovery by other scientists. Only then will the claimed discovery be agreed upon and the credits come your way.

The publication of papers is also becoming increasingly important in institutional contexts as a response to the growing culture of research assessment. An important factor in this is publishing in the relatively small number of journals that have, or are perceived to have, a high impact value. Lawrence (2003) has noted the distorting effect this can have on scientific publication.

In practice, authors can have many other reasons for writing and publishing papers. For example, in the preface to a book of his collected essays, the historian Cannadine

(2000, p. ix) has recorded some of the reasons why academic historians write essays and articles.

Historians write essays and articles for many specific and often unrelated reasons - to launch their careers, to establish a reputation, to keep their hand in; to please themselves, to impress their colleagues, to reach a broader audience; to sketch out a new idea, to anticipate a major work, to avoid writing a book; to take a break from a big project, to dabble but not delve too deeply, to revisit old friends and old haunts; to give as conference papers, to deliver as public lectures, to contribute to edited volumes; to indulge their scholarly curiosity, to make some (but not much) money, and (most recently and regrettably) to provide essential fodder for the Research Assessment Exercise.

While few of these - with the partial exception of the economic motive - would invalidate the possibility of depositing such essays in an institutional archive, their complexity might suggest that there may be little direct incentive for some researchers to do so.

5.2.3 Cultural differences between subject domains

The modern university contains a wide range of subject disciplines bundled together into a number of faculties or schools. While individual academics and researchers undertake many of the same types of activity, e.g. carrying out original research, publishing findings in papers and books, etc., subject disciplines often have their own distinctive culture. For example, Valauskas (1997) has noted how different the styles of communication, verification, debate and consensus can be amongst different academic disciplines. While these differences are most obvious between the broad categories of science, social science and the humanities, they also characterise subject divisions within these categories, e.g. between physics and chemistry (Ziman, 2000, p. 25). It may be no accident that those subject areas that had an existing pre-print culture (e.g., high-energy physics, computing science) or a tradition of issuing working papers (e.g., economics) have been among the most successful early-adopters of self-archiving and e-print repositories. While these differences may not matter over much within subject-based repositories, they may provide an organisational challenge to institutional ones. Researchers from disciplines that have a particularly high regard of peer-review (e.g., medicine or chemistry) may not be willing to use a repository that contains papers from other subject domains that have not been reviewed. The answer may be to have separate repositories for peer-reviewed and non-peer-reviewed papers or primarily to create institutional repositories at the department or faculty-level. The important thing to remember is that what may work for one subject discipline cannot just be assumed to be appropriate for others.

5.2.4 The diverse nature of research institutions

Many of the proponents of institutional repositories focus their arguments exclusively on universities and other academic institutions. So, for example, the position paper published by the Scholarly Publishing & Academic Resources Coalition (SPARC) defines institutional repositories as digital collections that capture and preserve the intellectual output of a single or multi-university community (Crow, 2002). However,

many research-active scientists work for other types of organisations, including hospitals, government agencies, commercial companies, museums, charities, etc. Bibliometric studies of UK research in the mid-1990s showed that while the majority of published research papers originated from educational institutions like universities, a significant proportion originated from other types of organisation (Hicks & Katz, 1996).

Medical institutions, industrial laboratories, research council and other government laboratories and non-profit institutes collectively seem to be as important as universities in the modern UK research system.
(<http://www.sussex.ac.uk/Users/sylvank/hickskatz/insttype.html>)

This raises the issue about what should happen to that proportion of published research that is not published by academic institutions.

The simple point is that institutional repositories organised solely by higher education institutions will exclude a significant amount of potential content. In this context, the value of a 'national service' that just gives access to the e-print output of just UK HE institutions looks unclear. Such a service would - at the least - need to work with other national services and repositories set up by non HE institutions.

6. Conclusions

This study of the prospects for institutional repositories in the UK shows that the current situation is uncertain. Great progress has been made in the development of standards and software tools that permit the easy creation of repositories. Chief amongst these has been the OAI-PMH, the University of Southampton's EPrints software and MIT's DSpace. The organisational side is less well developed, and some stakeholders have concerns about copyright, peer-review and long-term preservation. More seriously, the cultural dependence of academics on the existing journal system may mean that the take-up of self-archiving and other open-access methods may be incremental rather than rapid, focusing more on some subject disciplines than on others. This will be less than optimal for services like the national e-prints service proposed by ePrints UK. At the very least, ePrints UK should support the significant advocacy activity proposed by Pinfield (2003). Once sufficient content is available, it will be possible to evaluate both the proposed ePrints UK national service and the Web services designed to support their development and use.

Further papers in this series will cover some of the themes discussed here in more detail. These will cover business issues, collection development and research assessment.

Acknowledgements

UKOLN is funded by Resource: The Council for Museums, Archives and Libraries, the Joint Information Systems Committee (JISC) of the UK higher and further education funding councils, as well as by project funding from JISC, the European

Union and other organisations. UKOLN also receives support from the University of Bath, where it is based.

References

- Bide, M. (2002). *Open archives and intellectual property: incompatible world views?* Bath: Open Archives Forum, November. Available at: http://www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf
- Campbell, P. (2002). "Electronic futures in scientific communication and outreach." *Journal of Molecular Biology*, 319, 963-967.
- Cannadine, (2002). D. In *Churchill's shadow: confronting the past in modern Britain*. London: Allen Lane.
- Close, F. (1992). *Too hot to handle: the story of the race for cold fusion*. London: Penguin Books.
- Crow, R. (2002). *The case for institutional repositories: a SPARC position paper*. Washington, D.C.: Scholarly Publishing & Academic Resources Coalition. Available at: <http://www.arl.org/sparc/IR/ir.html>
- Fletcher, G. (2002). "Averting the crisis in medical publishing - open access journals." *Health Information on the Internet*, (30), 6-7.
- Gadd, E., Oppenheim, C. & Proberts, S. (2003). "The impact of copyright ownership on academic author self-archiving." *Journal of Documentation* (forthcoming) Available at: <http://www.lboro.ac.uk/departments/lis/disresearch/romeo/RoMEO%20Studies%201.pdf>
- Ginsparg, P. (1994). "First steps towards electronic research communication." *Computers in Physics*, 8(4), 390-396.
- Harnad, S. (2001). "The self-archiving initiative." *Nature*, 410, 1024-1025.
- Harnad, S. & Hey, J. (1995). "Esoteric knowledge: the scholar and scholarly publishing on the Net." In: Dempsey, L., Law, D. & Mowat, I. (eds.). *Networking and the future of libraries 2: managing the intellectual record*. London: Library Association Publishing, 110-116.
- Hicks, D. & Katz, J.S. (1996). "Systemic bibliometric indicators for the knowledge-based economy." OECD workshop on New Indicators for the Knowledge-Based Economy, Paris, 19-21 June 1996. Available at: <http://www.sussex.ac.uk/Users/sylvank/hickskatz/oecd.html>
- JISC Scholarly Communication Group. (2002). *Final report of the JISC Scholarly Communications Group (SCG) to the Research Support Libraries Group (RSLG)*. London: JISC. Available at: http://www.jisc.ac.uk/uploaded_documents/rslg.pdf
- LaPorte, R.E., Marler, E., Akazawa, S., Sauer, F., Gamboa, C., Shenton, C., Glosser, C., Villaseñor, A. & Maclure, M. (1995). "The Death of Biomedical Journals," *BMJ*, 310, 1387-1390.
- Lawrence, P.A. (2003). "The politics of publication." *Nature*, 422, 259-261.
- Lawrence, S. (2001). "Free online availability substantially increases a paper's impact." *Nature*, 411, 521.
- Lynch, C.A. (2003). "Institutional repositories: essential infrastructure for scholarship in the digital age." *ARL Bimonthly Report*, 226. Available at: <http://www.arl.org/newsltr/226/ir.html>
- Martin, R. (2003) "ePrints UK: creating a national e-print archive." *Ariadne*, 35, April. Available at: <http://www.ariadne.ac.uk/issue35/martin/>
- Meadows, A.J. (1991). "Scholarly communication and the serial." In: Brookfield, K. (ed.). *Scholarly communication and serials prices*. London: Bowker-Saur, 5-14.
- Nature. (2003). "Nature in 2003." *Nature*, 421, 1.

- Odlyzko, A. (1995) "Tragic loss or good riddance? The impending demise of traditional scholarly journals." *International Journal of Human-Computer Studies*, 42, 71-122. Available at: <http://www.dtc.umn.edu/~odlyzko/doc/tragic.loss.long.pdf>
- Odlyzko, A. (1997). "The economics of electronic journals." *First Monday*, 2(8), August. Available at: http://www.firstmonday.dk/issues/issue2_8/odlyzko/
- Peters, T.A. (2002). "Digital repositories: individual, discipline-based, institutional, consortial, or national?" *Journal of Academic Librarianship*, 28(6), 414-417.
- Pinfield, S. (2003). "Open archives and UK institutions: an overview." *D-Lib Magazine*, 9(3), March. Available at: <http://www.dlib.org/dlib/march03/pinfield/03pinfield.html>
- Powell, A. (2001). *An OAI approach to sharing subject gateway content*. WWW10: the Tenth International World Wide Web Conference, Hong Kong, 1-5 May 2001. Available at: <http://www10.org/cdrom/posters/1097.pdf>
- Roberts, R.J, Varmus, H.E., Ashburner, M., Brown, P.O., Eisen, M.B., Khosla, C., Kirchner, M., Nusse, R., Scott, M. & Wold, B. (2001). "Building a 'GenBank' of the published literature." *Science*, 291, 2318-2319.
- Rowland, F. (1997). "Print journals: fit for the future?" *Ariadne*, 7, 6-7. Available at: <http://www.ariadne.ac.uk/issue7/fytton/>
- Schaffner, A.C. (1994). "The future of scientific journals: lessons from the past." *Information Technology and Libraries*, 13(4), 239-247.
- Science. (2002). "Journal goes public." *Science*, 298, 2307.
- Smith, A. (2003) *New-model scholarship: how will it survive?* Washington, D.C.: Council on Library and Information Resources. Available at: <http://www.clir.org/pubs/abstract/pub114abst.html>
- Smith, M. (2002). "DSpace: an institutional repository from the MIT Libraries and Hewlett Packard Laboratories." In: M. Agosti & C. Thanos (eds.), *Research and advanced technology for digital libraries: 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002*. Lecture Notes in Computer Science, 2458. Berlin: Springer, 543-549.
- Smith, M, Barton, M., Bass, M., Branscholsky, M., McClellan, G., Stuve, D., Tansley, R. & Walker, J.H. (2003). "DSpace: an open source dynamic digital repository." *D-Lib Magazine*, 9(1), January. Available at: <http://www.dlib.org/dlib/january03/smith/01smith.html>
- Suber, P. (2003). "Removing the barriers to research: an introduction to open access for librarians." *College & Research Libraries News*, 64, 92-94, 113. Also available at: <http://www.earlham.edu/~peters/writing/acrl.htm>
- Valauskas, E.J. (1997). "Waiting for Thomas Kuhn: First Monday and the evolution of electronic journals." *First Monday*, 2(12), December. Available at: http://www.firstmonday.dk/issues/issue2_12/valauskas/
- Van de Sompel, H. & Lagoze, C. (2002). "Notes from the interoperability front: a progress report on the Open Archives Initiative" In: M. Agosti & C. Thanos (eds.), *Research and advanced technology for digital libraries: 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002*. Lecture Notes in Computer Science, 2458. Berlin: Springer, 144-157.
- Ziman, J.M. (1968). *Public knowledge: an essay concerning the social dimension of science*. Cambridge: Cambridge University Press.
- Ziman, J.M. (2000). *Real science: what it is, and what it means*. Cambridge: Cambridge University Press.